Elementary Indexes

April 29, 2020

Erwin Diewert,[1]
Discussion Paper 20-06,
Vancouver School of Economics,
The University of British Columbia,
Vancouver, Canada V6T 1L4.

Abstract

Elementary indexes are indexes that are used by national statistical agencies to construct a subindex of a national Consumer Price Index (CPI) at the first stage of aggregation. Usually, only price information is available to the National Statistical Office. The main elementary indexes that have been used over the years are the Carli, Dutot and Jevons indexes. These indexes are compared to each other using Taylor series approximations to the indexes. An axiomatic or test approach to elementary indexes is also described. Finally, Irving Fisher's rectification procedure that converts an elementary index into an index which satisfies the time reversal test is discussed.

Key Words: Superlative indexes, elementary indexes, the consumer price index, Fisher, Carli, Dutot and Jevons price indexes, the test approach to index number theory, Fisher's time rectification procedure.

Journal of Economic Literature Classification: C43, E31.

[1] University of British Columbia and University of New South Wales. Email: erwin.diewert@ubc.ca . This is a draft of Chapter 6 of *Consumer Price Index Theory*. This chapter draws heavily on Chapter 20 of the *Consumer Price Index Manual*; see ILO IMF OECD UNECE Eurostat World Bank (2004; 355-371). The author thanks Carsten Boldsen, Kevin Fox, Ronald Johnson, Claude Lamboray and Chihiro Shimizu for helpful comments.

1. Introduction

In all countries, the calculation of a Consumer Price Index proceeds in two (or more) stages. In the first stage of calculation, *elementary price indexes* are calculated for the *elementary expenditure aggregates* of a CPI. In the second and higher stages of aggregation, these elementary price indexes are combined to obtain higher level indexes using information on the expenditures on each elementary aggregate as weights. An *elementary aggregate* consists of the expenditures by a specified group of consumers on a relatively homogeneous set of products defined within the consumption classification used in the CPI.

At the first stage of aggregation, one of two possible situations can occur:

- Detailed price and quantity (or price and value) information on all transacted products in the elementary aggregate is available for the time period under consideration.[2]
- Only price information is available for the products in the aggregate under consideration. Moreover, the price information may be collected only for a *sample* of the entire set of product prices that are in scope.

At higher levels of aggregation, typically price and quantity (or value) information is available. Thus for higher levels of aggregation and for situations where detailed price and quantity information is available at the first stage of aggregation, the materials in previous chapters can be applied; i.e., Lowe, Laspeyres, Paasche and Fisher indexes can be used at higher levels of aggregation and at the elementary level if detailed price and quantity information is available. However, for situations where quantity or value information is not available, most of the index number theory outlined in previous chapters is not directly applicable. In this case, an elementary price index is a more primitive concept that relies on price data only. The situation where only price information is available will be the focus of this chapter.

The question of what is the most appropriate formula to use to construct an elementary price index is considered in this chapter.[3] The quality of a CPI depends heavily on the quality of the first stage of aggregation elementary indexes, which are the basic building blocks from which Consumer Price Indexes are constructed.

CPI compilers have to select *representative products* within an elementary aggregate and then collect a sample of prices for each of the representative products, usually from a sample of different outlets. The individual products whose prices are actually collected are described as the *sampled products*. Their prices are collected over successive time periods. An elementary price index is therefore typically calculated from two sets of *matched price observations*. In this chapter, we will assume that there are no missing observations and no changes in the quality of the products sampled so that the two sets of prices are perfectly matched. In the following chapter, we will consider alternative strategies when there are multiple time periods and missing observations; i.e., in chapter 7, we will discuss *multilateral index number theory*. In chapter 8, the treatment of new and disappearing goods and services and the related problems associated with measuring *quality change* will be discussed.

[2] With the increased availability of scanner data both for retail outlets as well as for individual consumers, the first situation is increasingly likely.

[3] The material in this chapter draws heavily on the contributions of Dalén (1992), Balk (1994) (2002) (2008) and Diewert (1995) (2002).

Before we define the elementary indexes used in practice, we will first consider in section 2 what is a suitable definition for an *ideal elementary index*. An ideal index will make use of expenditure data and so it cannot always be implemented in practice but it is useful to have an ideal target index. The problems involved in aggregating transaction prices for the same product over *time* are also discussed in this section. Thus the discussion in section 2 provides a theoretical target index for "practical" elementary price indexes that are constructed using only information on prices.

Section 3 provides some additional discussion about the problems involved in picking a suitable level of disaggregation for the elementary aggregates. Should the elementary aggregates have a regional dimension in addition to a product dimension? Should prices be collected from retail outlets or from households? These are the types of question discussed in this section.

Section 4 introduces the main elementary index formulae that are used in practice and section 5 develops some numerical relationships between the various "practical" indexes.

Section 6 develops the axiomatic or *test approach* to bilateral elementary indexes when only price information is available.

Section 7 contains material on the importance of the time reversal test.

Section 8 concludes with an overview of the various results.

2. Ideal Elementary Indexes

The aggregates covered by a CPI are usually arranged in the form of a tree like hierarchy (such as COICOP or NACE). An *aggregate* is a set of economic transactions pertaining to a set of commodities and a set of economic agents over a specified time period. Every economic transaction relates to the change of ownership of a specific, well defined commodity (good or service) at a particular place and date, and comes with a quantity and a price. A price index for an aggregate is typically calculated as a weighted average of the price indexes for the subaggregates, the (expenditure or sales) weights and type of average being determined by the index formula. One can descend in such a hierarchy as far as available information allows the weights to be decomposed. The lowest level aggregates are called *elementary* aggregates. They are basically of two types:

- Those for which all detailed price and quantity information is available;
- Those for which the statistician, considering the operational cost and the response burden of getting detailed price and quantity information about all the transactions, decides to make use of a representative sample of commodities and respondents.

As indicated above, the practical relevance of studying this topic is large. Since the elementary aggregates form the building blocks of a CPI, the choice of an inappropriate formula at this level can have a tremendous impact on the overall index.

In this section, it will be assumed that detailed price and quantity information for all transactions pertaining to the elementary aggregate for the two time periods under consideration is available. This assumption allows us to define an *ideal elementary aggregate*. Subsequent sections will relax this strong assumption about the availability of detailed price and quantity data on transactions but in any case, it is useful to have a theoretically ideal target for the "practical" elementary index.

The detailed price and quantity data, although perhaps not available to the statistician, is, in principle, available in the outside world. It is frequently the case that at the respondent level (i.e., at the outlet or firm level), some aggregation of the individual transactions information has been executed, usually in a form that suits the respondent's financial or management information system. This respondent determined level of information could be called the *basic information level*. This is, however, not necessarily the finest level of information that could be made available to the price statistician. One could always ask the respondent to provide more disaggregated information. For instance, instead of monthly data one could ask for weekly data; or, whenever appropriate, one could ask for regional instead of global data; or, one could ask for data according to a finer commodity classification. The only natural barrier to further disaggregation is the individual transaction level.[4]

It is now necessary to discuss a problem[5] that arises when detailed data on *individual transactions* are available, either at the level of the individual household or at the level of an individual outlet. Recall that in previous chapters, the price and quantity indexes, $P(p^0,p^1,q^0,q^1)$ and $Q(p^0,p^1,q^0,q^1)$, were introduced. These (bilateral) price and quantity indexes decomposed the value ratio V^1/V^0 into a price change part $P(p^0,p^1,q^0,q^1)$ and a quantity change part $Q(p^0,p^1,q^0,q^1)$. In this framework, it was taken for granted that the period t price and quantity for commodity i, p_i^t and q_i^t respectively, were well defined. However, these definitions are not straightforward since individual consumers may purchase the *same* item during period t *at different prices*. Similarly, if we look at the sales of a particular shop or outlet that sells to consumers, *the same item may sell at very different prices during the course of the period*. Hence before a traditional bilateral price index of the form $P(p^0,p^1,q^0,q^1)$ considered in previous chapters can be applied, there is a non-trivial *time aggregation problem* that must be solved in order to obtain the basic prices p_i^t and q_i^t that are the components of the price vectors p^0 and p^1 and the quantity vectors q^0 and q^1.

Walsh[6] and Davies (1924) (1932), suggested a solution to this time aggregation problem: the appropriate quantity at this very first stage of aggregation is the *total quantity purchased* of the narrowly defined item and the corresponding price is the value of purchases of this item divided by the total amount purchased, which is a *narrowly defined unit value*. In more recent times, most researchers have adopted the Walsh and Davies solution to the time aggregation problem.[7] Note that this solution to the time aggregation problem has the following advantages:

- The quantity aggregate is intuitively plausible, being the total quantity of the narrowly defined item purchased by the household (or sold by the outlet) during the time period under consideration;

[4] The material in this section is based on Balk (1994).

[5] This time aggregation problem was discussed briefly in chapter 2.

[6] Walsh explained his reasoning as follows: "Of all the prices reported of the same kind of article, the average to be drawn is the arithmetic; and the prices should be weighted according to the relative mass quantities that were sold at them." Correa Moylan Walsh (1901; 96). "Some nice questions arise as to whether only what is consumed in the country, or only what is produced in it, or both together are to be counted; and also there are difficulties as to the single price quotation that is to be given at each period to each commodity, since this, too, must be an average. Throughout the country during the period a commodity is not sold at one price, nor even at one wholesale price in its principal market. Various quantities of it are sold at different prices, and the full value is obtained by adding all the sums spent (at the same stage in its advance towards the consumer), and the average price is found by dividing the total sum (or the full value) by the total quantities." Correa Moylan Walsh (1921a; 88).

[7] See for example Szulc (1987; 13), Dalén (1992; 135), Reinsdorf (1994), Diewert (1995; 20-21), Reinsdorf and Moulton (1997) and Balk (2002).

- The product of the price times quantity equals the total value purchased by the household (or sold by the outlet) during the time period under consideration.

We will adopt this solution to the time aggregation problem as our concept for the price and quantity at this preliminary stage of aggregation.

Having decided on an appropriate theoretical definition of price and quantity for an item at the very lowest level of aggregation (i.e., a narrowly defined unit value and the total quantity sold of that item at the individual outlet), we now consider how to aggregate these narrowly defined elementary prices and quantities into an overall elementary aggregate. Suppose that there are N lowest level items or specific commodities in this chosen elementary category. Denote the period t quantity of item n by q_n^t and the corresponding time aggregated unit value price by p_n^t for t = 0,1 and for items n = 1,2,...,N. Define the period t quantity and price vectors as $q^t \equiv [q_1^t,q_2^t,...,q_N^t]$ and $p^t \equiv [p_1^t,p_2^t,...,p_N^t]$ for t = 0,1. It is now necessary to choose a theoretically ideal index number formula $P(p^0,p^1,q^0,q^1)$ that will aggregate the individual item prices into an overall aggregate price relative for the N items in the chosen elementary aggregate. However, this problem of choosing a functional form for $P(p^0,p^1,q^0,q^1)$ is *identical to the overall index number problem* that was addressed in previous chapters. In these previous chapters, four different approaches to index number theory were studied that led to specific index number formulae as being "best" from each perspective. From the viewpoint of *fixed basket approaches*, it was found that the Fisher (1922) and Walsh (1901) price indexes, P_F and P_W, appeared to be "best". From the viewpoint of the *test approach*, the Fisher index appeared to be "best". From the viewpoint of the *stochastic approach* to index number theory, the Törnqvist Theil index number formula P_T emerged as being "best". Finally, from the viewpoint of the *economic approach* to index number theory, the Walsh price index P_W, the Fisher ideal index P_F and the Törnqvist Theil index number formula P_T were all regarded as being equally desirable. It was also shown that the same three index number formulae numerically approximate each other very closely under certain conditions and so it will not matter very much which of these alternative indexes is chosen.[8] Hence, the *theoretically ideal elementary index number formula* is taken to be one of the three formulae $P_F(p^0,p^1,q^0,q^1)$, $P_W(p^0,p^1,q^0,q^1)$ or $P_T(p^0,p^1,q^0,q^1)$ where the period t quantity of item n, q_n^t, is the total quantity of that narrowly defined item purchased by the household during period t (or sold by the outlet during period t) and the corresponding price for item n is p_n^t, the time aggregated unit value, for t = 0,1 and for items n = 1,2,...,N.

In the following sections, various "practical" elementary price indexes will be defined. These indexes do not have quantity weights and thus are functions only of the price vectors p^0 and p^1. Thus when a practical elementary index number formula, say $P_E(p^0,p^1)$, is compared to an ideal elementary price index, say the Fisher price index $P_F(p^0,p^1,q^0,q^1)$, then obviously P_E will differ from P_F because the *prices are not weighted according to their economic importance* in the practical elementary formula. It is useful to list the following possible sources of difference between a practical elementary price index $P_E(p^0,p^1)$ and an ideal target index:

- *Weighting bias* or more generally, *formula bias*; i.e., a price index of the form $P_E(p^0,p^1)$ is not able to weight prices according to the economic importance of the product in the consumer's total expenditures on the group of products under consideration.[9]

[8] Theorem 5 in Diewert (1978; 888) showed that P_F, P_T and P_W will approximate each other to the second order around an equal price and quantity point. However, if there are violent fluctuations in prices and quantities, a second order approximation to any one of these formulae may not be very accurate.

[9] For materials on how to measure formula bias, see Diewert (1998), White (1999) (2000) and chapter 7.

- *Sampling bias*; i.e., the statistical agency may not be able to collect information on all N products in the elementary aggregate; i.e., only a sample of the N prices may be collected.[10]
- *Time aggregation bias*; i.e., even if a price for a narrowly defined item is collected by the statistical agency, it may not be equal to the theoretically appropriate time aggregated unit value price.[11]
- *Item aggregation bias* or *unit value bias*. The statistical agency may classify certain distinct products as being essentially equivalent and thus the unit value aggregate for this group of aggregated products may not take into account possible significant quality differences in the group of aggregated products. For example, products that are thought to be very similar and are sold in the same units of measurement could be treated as a single product.[12]
- *Aggregation over agents* or *aggregation over entities bias* or *aggregation over outlets bias*. The unit value for a particular item may be constructed by aggregating over all households in a region or a certain demographic class or by aggregating over all outlets or shops that sell the item in a particular region.[13]
- *New and disappearing products bias*; i.e., $P_E(p^0,p^1)$ measures price change only over matched products for the two periods being compared; new products and disappearing products are ignored in standard elementary indexes that depend only on prices.[14]

Approximations to the numerical differences between various elementary indexes of the form $P_E(p^0,p^1)$ and various superlative indexes will be developed in chapter 7.

In the following section, the problems of aggregation and classification will be discussed in more detail.

3. Aggregation and Classification Problems for Elementary Aggregates

Hawkes and Piotrowski (2003) noted that the definition of an elementary aggregate involves aggregation over *four* possible dimensions:[15]

- A *time* dimension; i.e., the item unit value could be calculated for all item transactions for a year, a month, a week, or a day.

---

[10] This is a specialized topic with a long history. It will not be covered in this volume.

[11] Many statistical agencies send price collectors to various outlets on certain days of the month to collect list prices of individual items. Usually, price collectors do not work on weekends when many sales take place and thus the collected prices may not be fully representative of all transactions that occur. Thus these collected prices can be regarded only as approximations to the time aggregated unit values for those items.

[12] For materials on unit value bias, see Diewert and von der Lippe (2010) and Silver (2010) (2011) and the additional references in these papers.

[13] For materials on possible methods to measure outlet substitution bias, see Diewert (1998). The problems associated with measuring aggregation over consumers bias were noted in the final sections of chapter 5.

[14] This problem was addressed in section 14 of chapter 5. It will be addressed in more detail in chapters 7 and 8.

[15] Hawkes and Piotrowski (2003; 31) combined the spatial and sectoral dimensions into the spatial dimension. They also acknowledged the pioneering work of Theil (1954), who identified three dimensions of aggregation: aggregation over individuals, aggregation over commodities and aggregation over time. It should be noted that William Hawkes was a pioneer in realizing the importance of scanner data for the construction of Consumer Price Indexes; see Hawkes (1997). Other important contributors include Reinsdorf (1996), Silver (1995), Silver and Heravi (2001) (2003) (2005), de Haan and van der Grient (2011), Ivancic, Diewert and Fox (2011) and de Haan and Krsinich (2014).

- A *spatial* dimension; i.e., the item unit value could be calculated for all item transactions in the country, province or state, city, neighbourhood, or individual location.
- A *product* dimension; i.e., the item unit value could be calculated for all item transactions in a broad general category (e.g., food), in a more specific category (e.g., margarine), for a particular brand (ignoring package size) or for a particular narrowly defined item (e.g., a particular AC Nielsen universal product code).
- A *sectoral* (or *entity* or *economic agent*) dimension; i.e., the item unit value could be calculated for a particular class of households or a particular class of outlets.

Each of the above dimensions for choosing the domain of definition for an elementary aggregate will be discussed in turn.

As the time period is compressed, several problems emerge:

- Purchases (by households) and sales (by outlets) become erratic and sporadic. Thus the frequency of unmatched purchases or sales from one period to the next increases and in the limit (choose the time period to be one minute), nothing will be matched and bilateral index number theory fails at the individual consumer level.[16]
- As the time period becomes shorter, chained indexes exhibit more "drift"; i.e., if the data at the end of a chain of periods reverts to the data in the initial period, the chained index does not revert back to unity. As was discussed in section 8 of chapter 2, it is only appropriate to use chained indexes when the underlying price and quantity data exhibit relatively smooth trends. When the time period is short, seasonal fluctuations[17] and periodic sales and advertising campaigns[18] can cause prices and quantities to oscillate (or "bounce" to use Szulc's (1983; 548) term) and hence it is not appropriate to use chained indexes under these circumstances. If fixed base indexes are used in this short time period situation, then the results will usually depend very strongly on the choice of the base period. In the seasonal context, not all commodities may even be in the marketplace during the chosen base period.[19] All of these problems can be mitigated by choosing a longer time period so that trends in the data will tend to dominate the short term fluctuations.
- As the time period contracts, virtually all goods become *durable* in the sense that they yield services not only for the period of purchase but for subsequent periods. Thus the period of purchase or acquisition becomes different from the periods of use, leading to many complications.[20]
- As the time period contracts, users will usually not be particularly interested in the short term fluctuations of the resulting index and there will be demands for smoothing the

[16] This problem was noted in section 19 of chapter 5. David Richardson (2003; 51) also made this point: "Defining items with a finer granularity, as is the case if quotes in different weeks are treated as separate items, results in more missing data and more imputations." However, high frequency consumer price indexes could be successfully constructed if aggregation over households or outlets is permitted.
[17] See chapter 11 below for a monthly seasonal example where chained month to month indexes are a disaster.
[18] See Feenstra and Shapiro (2003) for an example of a weekly superlative index that exhibits massive chain drift. Substantial chain drift can also occur using monthly indexes; see Szulc (1983) (1987). See Richardson (2003; 50-51) and Ivancic, Diewert and Fox (2011) for additional discussions of the issues involved in choosing weekly unit values versus monthly unit values.
[19] See chapter 9 below for suggested solutions to these seasonality problems.
[20] See chapter 10 below for more material on the possible CPI treatment of durable goods.

> necessarily erratic results. Put another way, users will desire a way of summarizing the weekly or daily movements in the index into monthly or quarterly movements in prices. Hence from the viewpoint of meeting the needs of users, there may be relatively little demand for high frequency indexes.

In view of the above considerations, it is recommended that the index number time period be at least four consecutive weeks or a month.[21]

It is also necessary to choose the spatial dimension of the elementary aggregate. Should item prices in each city or region be considered as separate aggregates or should a national item aggregate be constructed? Obviously, if it is desired to have regional CPIs that aggregate up to a national CPI, then it will be necessary to collect item prices by region. However, it is not clear how fine the “regions” should be. It could be as fine as a grouping of households in a postal code or to individual outlets across the country.[22] There does not seem to be a clear consensus on what the optimal degree of spatial disaggregation should be.[23] Each statistical agency will have to make its own judgements on this matter, taking into account the costs of data collection and the demands of users for a spatial dimension for the CPI.

How detailed should the product dimension be? The possibilities range from regarding all commodities in a general category as being equivalent to the other extreme, where only a commodity in a particular package size made by a particular manufacturer or service provider is regarded as being equivalent. All things being equal, Triplett (2004) stressed the advantages of matching products at the most detailed level possible, since this will prevent quality differences from clouding the period to period price comparisons. This is sensible advice but then what are the drawbacks to working with the finest possible commodity classification? The major drawback is that the finer the classification is, the more difficult it will be to *match* the item purchased or sold in the base period to the same item in the current period. Hence, the finer the product classification, the smaller will be the number of matched price comparisons that are possible.[24] This would not be a problem if the unmatched prices followed the same trend as the matched

[21] If there is very high inflation in the economy (or even hyperinflation), then it may be necessary to move to weekly or even daily indexes. Also, it should be noted that some index number theorists feel that new theories of consumer behavior should be developed that could utilize weekly or daily data: “Some studies have endorsed unit values to reduce high frequency price variation, but this implicitly assumes that the high frequency variation represents simply noise in the data and is not meaningful in the context of a COLI. That is debatable. We need to develop a theory that confronts the data, not truncate the data to fit the theory.” Jack E. Triplett (2003; 153). However, until such new theories are adequately developed, it seems pragmatic to define the item unit values over months or quarters rather than days or weeks.

[22] Iceland no longer uses regional weights but uses individual outlets as the primary geographical unit; see Gudnason (2003; 18).

[23] Hawkes and Piotrowski note that it is quite acceptable to use national elementary aggregates when making international comparisons between countries: “When we try to compare egg prices across geography, however, we find that lacing across outlets won’t work, because the eyelets on one side of the shoe (or outlets on one side of the river) don’t match up with those on the other side. Thus, in making interspatial comparisons, we have no choice but to aggregate outlets all the way up to the regional (or, in the case of purchasing power parities, national) level. We have no hesitation about doing this for interspatial comparisons, but we are reluctant to do so for intertemporal ones. Why is this?” William J. Hawkes and Frank W. Piotrowski (2003; 31-32). An answer to their question is that it is preferable to match like with like as closely as possible which leads statisticians to prefer the finest possible level of aggregation, which, in the case of intertemporal comparisons, would be the individual household or the individual outlet. However, in making cross region comparisons, matching is not possible unless regional item aggregates are formed, as Hawkes and Piotrowski point out above.

[24] This is part of the *matching problem* discussed at the end of chapter 5.

ones in a particular elementary aggregate, but in at least some circumstances, this will not be the case.[25] Thus the finer the classification system is, the more work (in principle) there will be for the statistical agency to quality adjust or impute the prices that do not match. Choosing a relatively coarse classification system leads to a very cost efficient system of quality adjustment (i.e., essentially no explicit quality adjustment or imputation is done for the prices that do not exactly match) but it may not be very accurate. Thus all things considered, it seems preferable to choose the finest possible classification system.

The final issue in choosing a classification scheme is the issue of choosing a sectoral dimension; i.e., should the unit value for a particular item be calculated for a particular outlet or a particular household or for a class of outlets or households? Before this question can be answered, it is necessary to ask whether the individual outlet or the individual household is the appropriate finest level of entity classification. If the economic approach to the consumer price index is taken, then the individual household is the appropriate finest level of entity classification.[26] Obviously, a single household will not work very well as the basic unit of entity observation due to the sporadic nature of many purchases by an individual household; i.e., there will be tremendous difficulties in matching prices across periods for individual households. However, for a grouping of "similar" households that is sufficiently large, it does become feasible in theory to use the grouped household as the entity classification rather than the outlet as is usually done. This is not usually done because of the costs and difficulties involved in collecting individual household data on prices and expenditures.[27] Thus price information is usually collected from retail establishments or outlets that sell mainly to households. Matching problems are mitigated using this strategy (but not eliminated) because the retail outlet generally sells the same items on a continuing basis.

If expenditures by all households in a region are aggregated together, will they equal sales by the retail outlets in the region? Under certain conditions, the answer to this question is yes. The conditions are that the outlets do not sell any items to purchasers who are not local households (no regional exports or sales to local businesses or governments) and that the regional households do not make any purchases of consumption items other than from the local outlets (no household imports or transfers of commodities to local households by governments). Obviously, these restrictive conditions will not be met in practice but they may hold as a first approximation.

The effects of *regional aggregation* and *product aggregation* can be examined, thanks to a study by Koskimäki and Ylä-Jarkko (2003). This study utilized scanner data for the last week in September 1998 and September 2000 on butter, margarine and other vegetable fats, vegetable

[25] Silver and Heravi (2001) (2003; 286) (2005) and Koskimäki and Vartia (2001) stressed this point and presented empirical evidence to back up their point. Feenstra (1994) and Balk (2000) used the assumption of CES preferences to deal with the new products problem. Their approaches will be discussed in chapter 8.

[26] This point has been made emphatically by two authors in a book on scanner data and price indexes: "In any case, unit values across stores are not the prices actually faced by households and do not represent the per period price in the COLI, even if the unit values are grouped by type of retail outlet." Jack E. Triplett (2003; 153-154). "Furthermore, note that the relationship being estimated is not a proper consumer demand function but rather an 'establishment sales function'. Only after making further assumptions – for example, fixing the distribution of consumers across establishments – is it permissible to jump to demand functions". Eduardo Ley (2003; 380).

[27] However, it is possible to collect accurate household data in certain circumstances; see Gudnason (2003), who pioneered a receipts methodology for collecting household price and expenditure data in Iceland. Also, in the future, as monetary transactions are replaced by debit and credit card transactions, it will become possible to construct individual household estimates of real consumption, provided that product codes are included in the transaction records.

oils, soft drinks, fruit juices and detergents; this information was provided by the AC Nielsen company for Finland. At the finest level of item classification (the AC Nielsen Universal Product Code), the number of individual items in the sample was 1028. The total number of outlets in the sample was 338. Koskimäki and Ylä-Jarkko considered four levels of spatial disaggregation:

- The entire country (1 level);
- Provinces (4 levels);
- AC Nielsen regions (15 levels);
- Individual outlets (338 levels).

They also considered four levels of product disaggregation:

- The COICOP 5 digit classification (6 levels);
- The COICOP 7 digit classification (26 levels);
- The AC Nielsen brand classification (266 levels);
- The AC Nielsen individual Universal Product Code (1028 distinct products).

In order to illustrate the ability to match products over the two year period as a function of the degree of fineness of the classification, Koskimäki and Ylä-Jarkko (2003; 10) presented a table that shows that the proportion of transactions that could be matched across the two years fell steadily as the fineness of the classification scheme increased. At the highest level of aggregation (the national and COICOP 5 digit), all transactions could be matched over the two year period but at the finest level of aggregation (338 outlets times 1028 individual products or 347,464 classification cells in all), only 61.7 % of the value of transactions in 2000 could be matched back to their 1998 counterparts. Their Table 7 is reproduced as Table 1 below.

Table 1: Proportion of Transactions in 2000 that Could be Matched to 1998

	COICOP 5 digit	**COICOP 7 digit**	**AC Nielsen Brand**	**AC Nielsen UPC**
Country	1.000	1.000	.982	.801
Province	1.000	1.000	.975	.774
AC Nielsen Region	1.000	1.000	.969	.755
Individual Outlet	.904	.904	.846	.617

For each of the above 16 levels of product and regional disaggregation, for the products that were available in September of 1998 and 2000, Koskimäki and Ylä-Jarkko (2003; 9) calculated Laspeyres and Fisher price indexes. They found substantial differences in these indexes as the degree of disaggregation increased.

Another study on the effects of alternative methods of unit value aggregation over outlets (i.e., treat each unit value for each product in each store as a unique product versus aggregating products over stores and chains) was undertaken by Ivancic and Fox (2013). They used 65 weeks of scanner data on the sales of different types of instant coffee sold by four supermarket chains in Australia in 110 stores where the data were collected between February 1997 and April 1998. It contains information on 110 stores which belong to four supermarket chains located in the metropolitan area of one of the major capital cites in Australia. These stores accounted for over 80 percent of grocery sales in the various capital cities of Australia during this period. expenditure for this particular item category. After data exclusions 436,103 weekly observations on

157 coffee items were used in their study.[28] Their results on alternative methods of aggregation can be summarized as follows:

"The results show that when non-superlative index numbers are used to calculate price change, aggregation choices can have a huge impact. However, the issue of aggregation seems to become relatively trivial when the standard Fisher and Törnqvist superlative indexes are used, with an extremely close range of estimates of price change found across different aggregation methods. This result seems to provide further support for the use of these superlative indexes over the use of non-superlative indexes to estimate price change." Lorraine Ivancic and Kevin J. Fox (2013; 643).

The non-superlative index numbers[29] were chained Laspeyres and Paasche indexes and the superlative indexes were chained Fisher and Törnqvist Theil indexes.

Thus the problem of determining the "best" unit value to insert into an index number formula is far from settled. We will look at this problem again in Chapter 11.

Another issue that arises in the context of defining exactly what prices and quantities should be entered into an index number formula. Some statistical agencies exclude sale prices. In general, this is not a recommended practice. Fox and Syed (2016; 404) found that the exclusion of sales prices can introduce a substantial bias. They also found that even when sales prices are included they are systematically under-weighted, but the under-weighting remains fairly stable over time so that inflation measurement is not significantly affected. They also found evidence that the typical practice of using data from an incomplete period in constructing unit values can lead to an upward bias in the resulting price index.[30]

4. Some Elementary Indexes that Have Been Suggested Over the Years

Suppose that there are N commodities in a chosen elementary category. Denote the period t price of item n by p_n^t for t = 0,1 and for items n = 1,2,...,N. As usual, define the period t price vector as $p^t \equiv [p_1^t,p_2^t,...,p_N^t]$ for t = 0,1.

The first simple elementary index number formula is due to the French economist Dutot (1738):

$$(1)\quad P_D(p^0,p^1) \equiv [\sum_{n=1}^N (1/N)\, p_n^1]/[\sum_{n=1}^N (1/N)\, p_n^0] = [\sum_{n=1}^N p_n^1]/[\sum_{n=1}^N p_n^0] = p^1\cdot 1_N/p^0\cdot 1_N.$$

Thus the Dutot elementary price index is equal to the arithmetic average of the N period 1 prices divided by the arithmetic average of the N period 0 prices.

The second simple elementary index number formula is due to the Italian economist Carli (1764):

$$(2)\quad P_C(p^0,p^1) \equiv \sum_{n=1}^N (1/N)(p_n^1/p_n^0).$$

Thus the Carli elementary price index is equal to the *arithmetic* average of the N item price ratios or price relatives, p_n^1/p_n^0. This formula was already encountered in our study of the unweighted stochastic approach to index numbers; recall definition (2) in chapter 4 above.

[28] Their paper also lists some related studies.
[29] The weekly unit values by product were aggregated into monthly unit values.
[30] Diewert, Fox and de Haan (2016) also found this effect.

The third simple elementary index number formula is due to the English economist Jevons (1865):

(3) $P_J(p^0,p^1) \equiv \prod_{n=1}^{N} (p_n^1/p_n^0)^{1/N}$.

Thus the Jevons elementary price index is equal to the *geometric* average of the N item price ratios or price relatives, p_n^1/p_n^0. Again, this formula was introduced as formula (4) in our discussion of the unweighted stochastic approach to index number theory in chapter 4 above.

The fourth elementary index number formula P_H is the *harmonic* average of the N item price relatives and it was first suggested in passing as an index number formula by Jevons (1865; 121) and Coggeshall (1887):

(4) $P_H(p^0,p^1) \equiv [\sum_{n=1}^{N} (1/N)(p_n^1/p_n^0)^{-1}]^{-1}$.

Finally, the fifth elementary index number formula is the geometric average of the Carli and harmonic formulae; i.e., it is *the geometric mean of the arithmetic and harmonic means* of the N price relatives:

(5) $P_{CSWD}(p^0,p^1) \equiv [P_C(p^0,p^1)\ P_H(p^0,p^1)]^{1/2}$.

This index number formula was first suggested by Fisher (1922; 472) as his formula 101. Fisher also observed that, empirically for his data set, P_{CSWD} was very close to the Jevons index, P_J, and these two indexes were his "best' unweighted index number formulae. In more recent times, Carruthers, Sellwood and Ward (1980; 25) and Dalén (1992; 140) also proposed P_{CSWD} as an elementary index number formula.

It should be noted that the Jevons index is now the most commonly used elementary index (when only price information is available). The Dutot and Carli formulae are used by a few statistical agencies.

Having defined the most commonly used elementary formulae, the question now arises: which formula is "best"? Obviously, this question cannot be answered until desirable properties for elementary indexes are developed. This will be done in a systematic manner in section 6 below (using the test approach) but in the present section, one desirable property for an elementary index will be noted. This is the *time reversal test*, which was noted earlier in chapters 2 and 3. In the present context, this test for the elementary index $P(p^0,p^1)$ becomes:

(6) $P(p^0,p^1)P(p^1,p^0) = 1$.

This test says that if the prices in period 2 revert to the initial prices of period 0, then the product of the price change going from period 0 to 1, $P(p^0,p^1)$, times the price change going from period 1 to 2, $P(p^1,p^0)$, should equal unity; i.e., under the stated conditions, we should end up where we started.[31] It can be verified that the Dutot, Jevons and Carruthers, Sellwood, Ward and Dalén indexes, P_D, P_J and P_{CSWD}, all satisfy the time reversal test but that the Carli and Harmonic indexes, P_C and P_H, fail this test. In fact, these last two indexes fail the test in the following *biased* manner:

[31] This test can also be viewed as a special case of Walsh's (1901) Multiperiod Identity Test, (63) in chapter 2.

(7) $P_C(p^0,p^1)\, P_C(p^1,p^0) \geq 1$;
(8) $P_H(p^0,p^1)\, P_H(p^1,p^0) \leq 1$

with strict inequalities holding in (7) and (8) provided that the period 1 price vector p^1 is not proportional to the period 0 price vector p^0.[32] Thus the Carli index will generally have an *upward bias* while the Harmonic index will generally have a *downward bias*. Fisher (1922; 66 and 383) was quite definite in his condemnation of the Carli index due to its upward bias.[33] Because it fails the time reversal test, the Carli index should not be used in compiling elementary price indexes for the Harmonized Index of Consumer Prices (HICP) that is the official Eurostat index used to compare consumer prices across European Union countries.

In the following section, some numerical relationships between the five elementary indexes defined in this section will be established. Then in the subsequent section, a more comprehensive list of desirable properties for elementary indexes will be developed and the five elementary formulae will be evaluated in the light of these properties or tests.

5. Numerical Relationships between Some Elementary Indexes

It can be shown[34] that the Carli, Jevons and Harmonic elementary price indexes satisfy the following inequalities:

(9) $P_H(p^0,p^1) \leq P_J(p^0,p^1) \leq P_C(p^0,p^1)$;

i.e., the Harmonic index is always equal to or less than the Jevons index, which in turn, is always equal to or less than the Carli index. In fact, the strict inequalities in (9) will hold provided that the period 0 vector of prices, p^0, is not proportional to the period 1 vector of prices, p^1.

The inequalities (9) do not tell us by how much the Carli index will exceed the Jevons index and by how much the Jevons index will exceed the Harmonic index. Hence, in the remainder of this section, some approximate relationships between the five indexes defined in the previous section will be developed that will provide some practical guidance on the relative magnitudes of each of the indexes.

The first approximate relationship that will be derived is between the Jevons index P_J and the Dutot index P_D. For each period t, define the *arithmetic mean* of the N prices pertaining to that period as follows:

(10) $p^{t*} \equiv \sum_{n=1}^{N} (1/N)\, p_n^t$; $\qquad t = 0,1.$

Now define (implicitly) the *multiplicative deviation* of the nth price in period t relative to the mean price in that period, e_n^t, as follows:

(11) $p_n^t = p^{t*}(1+e_n^t)$; $\qquad n = 1,...,N ;\ t = 0,1.$

[32] These inequalities follow from the fact that a harmonic mean of N positive numbers is always equal to or less than the corresponding arithmetic mean; see Walsh (1901;517) or Fisher (1922; 383-384). This inequality is a special case of Schlömilch's Inequality; see Hardy, Littlewood and Pólya (1934; 26).
[33] See also Szulc (1987; 12) and Dalén (1992; 139). Dalén (1994; 150-151) provided some nice intuitive explanations for the upward bias of the Carli index.
[34] Each of the three indexes P_H, P_J and P_C is a mean of order r where r equals −1, 0 and 1 respectively and so the inequalities follow from Schlömilch's inequality.

Note that (10) and (11) imply that the deviations e_n^t sum to zero in each period; i.e., we have:

$$(12)\ \sum_{n=1}^{N} e_n^t = 0\ ; \qquad t = 0,1.$$

Note that the Dutot index can be written as the ratio of the mean prices, p^{1*}/p^{0*}; i.e., we have:

$$(13)\ P_D(p^0,p^1) = p^{1*}/p^{0*}.$$

Now substitute equations (11) into the definition of the Jevons index, (3):

$$\begin{aligned}(14)\ P_J(p^0,p^1) &= \prod_{n=1}^{N} [p^{1*}(1+e_n^1)/p^{0*}(1+e_n^0)]^{1/N} \\ &= [p^{1*}/p^{0*}] \prod_{n=1}^{N} [(1+e_n^1)/(1+e_n^0)]^{1/N} \\ &= P_D(p^0,p^1)\ f(e^0,e^1) \qquad \text{using definition (1)}\end{aligned}$$

where $e^t \equiv [e_1^t,...,e_N^t]$ for t = 0 and 1, and the function f is defined as follows:

$$(15)\ f(e^0,e^1) \equiv \prod_{N=1}^{N} [(1+e_n^1)/(1+e_n^0)]^{1/N}.$$

Expand $f(e^0,e^1)$ by a second order Taylor series approximation around $e^0 = 0_N$ and $e^1 = 0_N$. Using (12), it can be verified[35] that we obtain the following second order approximate relationship between P_J and P_D:

$$\begin{aligned}(16)\ P_J(p^0,p^1) &\approx P_D(p^0,p^1)[1 + (1/2N)e^0\cdot e^0 - (1/2N)e^1\cdot e^1] \\ &= P_D(p^0,p^1)[1 + (1/2)\mathrm{var}(e^0) - (1/2)\mathrm{var}(e^1)]\end{aligned}$$

where $\mathrm{var}(e^t)$ is the variance of the period t multiplicative deviations; i.e., for t = 0,1:

$$\begin{aligned}(17)\ \mathrm{var}(e^t) &\equiv (1/N)\sum_{n=1}^{N} (e_n^t - e^{t*})^2 \\ &= (1/N)\sum_{n=1}^{N} (e_n^t)^2 \qquad \text{since } e^{t*} = 0 \text{ using (12)} \\ &= (1/N)e^t\cdot e^t\ .\end{aligned}$$

Under normal conditions[36], the variance of the deviations of the prices from their means in each period is likely to be approximately constant and so under these conditions, the Jevons price index will approximate the Dutot price index to the second order.

Note that with the exception of the Dutot formula, the remaining four elementary indexes defined in section 4 are functions of the relative prices of the N items being aggregated. This fact is used in order to derive some approximate relationships between these four elementary indexes. Thus define the *nth price relative* as

$$(18)\ r_n \equiv p_n^1/p_n^0\ ; \qquad n = 1,...,N.$$

Define the *arithmetic mean of the n price relatives* as

[35] This approximate relationship was first obtained by Carruthers, Sellwood and Ward (1980; 25).

[36] If there are significant changes in the overall inflation rate, some studies indicate that the variance of deviations of prices from their means can also change. Also if N is small, then there will be sampling fluctuations in the variances of the prices from period to period, leading to random differences between the Dutot and Jevons indexes.

(19) $r^* \equiv (1/N)\sum_{n=1}^{N} r_n = P_C(p^0,p^1)$

where the last equality follows from the definition (2) for the Carli index. Finally, define (implicitly) the *deviation* e_n of the nth price relative r_n from the arithmetic average of the N price relatives r^* as follows:

(20) $r_n = r^*(1+e_n)$; $\quad n = 1,...,N.$

Note that (19) and (20) imply that the deviations e_n sum to zero; i.e., we have:

(21) $\sum_{n=1}^{N} e_n = 0.$

Now substitute equations (20) into the definitions of P_C, P_J, P_H and P_{CSWD}, (2)-(5) above, in order to obtain the following representations for these indexes in terms of the vector of deviations, $e \equiv [e_1,...,e_N]$:[37]

(22) $P_C(p^0,p^1) = \sum_{n=1}^{N} (1/N)r_n = r^* \equiv r^*f_C(e)$;
(23) $P_J(p^0,p^1) = \prod_{n=1}^{N} r_n^{1/N} = r^*\prod_{n=1}^{N} (1+e_n)^{1/N} \equiv r^*f_J(e)$;
(24) $P_H(p^0,p^1) = [\sum_{n=1}^{N} (1/N)(r_n)^{-1}]^{-1} = r^*[\sum_{n=1}^{N} (1/N)(1+e_n)^{-1}]^{-1} \equiv r^*f_H(e)$;
(25) $P_{CSWD}(p^0,p^1) = [P_C(p^0,p^1)P_H(p^0,p^1)]^{1/2} = r^*[f_C(e)f_H(e)]^{1/2} \equiv r^*f_{CSWD}(e)$

where the last equation in (22)-(25) serves to define the deviation functions, $f_C(e)$, $f_J(e)$, $f_H(e)$ and $f_{CSWD}(e)$. The second order Taylor series approximations to each of these functions[38] around the point $e = 0_N$ are:

(26) $f_C(e) \approx 1$;
(27) $f_J(e) \approx 1 - (1/2N)e{\cdot}e = 1 - (1/2)\mathrm{var}(e)$;
(28) $f_H(e) \approx 1 - (1/N)e{\cdot}e = 1 - \mathrm{var}(e)$;
(29) $f_{CSWD}(e) \approx 1 - (1/2N)e{\cdot}e = 1 - (1/2)\mathrm{var}(e)$

where we have made repeated use of (21) in deriving the above approximations.[39] Thus to the second order, the Carli index P_C will *exceed* the Jevons and Carruthers Sellwood Ward Dalén indexes, P_J and P_{CSWD}, by $(1/2)r^*\mathrm{var}(e)$, which is r^* times one half the variance of the N price relatives p_n^1/p_n^0. Similarly, to the second order, the Harmonic index P_H will *lie below* the Jevons and Carruthers Sellwood Ward Dalén indexes, P_J and P_{CSWD}, by r^* times one half the variance of the N price relatives p_n^1/p_n^0.

Thus empirically, it is expected that the Jevons and Carruthers Sellwood Ward and Dalén indexes will be very close to each other. Using the previous approximation result (16), it is expected that the Dutot index P_D will also be fairly close to P_J and P_{CSWD}, with some fluctuations over time due to changing variances of the period 0 and 1 deviation vectors, e^0 and e^1. Thus it is expected that these three elementary indexes will give much the same numerical answers in empirical applications. On the other hand, the Carli index can be expected to be substantially *above* these three indexes, with the degree of divergence growing as the variance of the N price relatives

[37] Note that the vector of deviations e defined by equations (20) is different from the deviation vectors e^0 and e^1 defined by equations (11).
[38] From (22), it can be seen that $f_C(e)$ is identically equal to 1 so that (26) will be an exact equality rather than an approximation.
[39] These second order approximations are due to Dalén (1992; 143) for the case $r^* = 1$ and to Diewert (1995; 29) for the case of a general r^*.

grows. Similarly, the Harmonic index can be expected to be substantially *below* the three middle indexes, with the degree of divergence growing as the variance of the N price relatives grows.

6. The Test Approach to Elementary Indexes

Recall that in chapter 3, the axiomatic approach to bilateral price indexes $P(p^0,p^1,q^0,q^1)$ was developed. In the present section, the elementary price index $P(p^0,p^1)$ depends only on the period 0 and 1 price vectors, p^0 and p^1 respectively, so that the elementary price index does not depend on the period 0 and 1 quantity vectors, q^0 and q^1. One approach to obtaining new tests or axioms for an elementary index is to look at the twenty or so axioms that were listed in Chapter 3 for bilateral price indexes $P(p^0,p^1,q^0,q^1)$ and adapt those axioms to the present context; i.e., use the old bilateral tests for $P(p^0,p^1,q^0,q^1)$ that do not depend on the quantity vectors q^0 and q^1 as tests for an elementary index $P(p^0,p^1)$.[40] This approach will be utilized in the present section.

The first eight tests or axioms are reasonably straightforward and uncontroversial:

T1: *Continuity*: $P(p^0,p^1)$ is a continuous function of the N positive period 0 prices $p^0 \equiv [p_1^0,...,p_N^0]$ and the N positive period 1 prices $p^1 \equiv [p_1^1,...,p_N^1]$.

T2: *Identity*: $P(p,p) = 1$; i.e., the period 0 price vector equals the period 1 price vector, then the index is equal to unity.

T3: *Monotonicity in Current Period Prices*: $P(p^0,p^1) < P(p^0,p)$ if $p^1 < p$; i.e., if any period 1 price increases, then the price index increases.

T4: *Monotonicity in Base Period Prices*: $P(p^0,p^1) > P(p,p^1)$ if $p^0 < p$; i.e., if any period 0 price increases, then the price index decreases.

T5: *Proportionality in Current Period Prices*: $P(p^0,\lambda p^1) = \lambda P(p^0,p^1)$ if $\lambda > 0$; i.e., if all period 1 prices are multiplied by the positive number λ, then the initial price index is also multiplied by λ.

T6: *Inverse Proportionality in Base Period Prices*: $P(\lambda p^0,p^1) = \lambda^{-1} P(p^0,p^1)$ if $\lambda > 0$; i.e., if all period 0 prices are multiplied by the positive number λ, then the initial price index is multiplied by $1/\lambda$.

T7: *Mean Value Test*: $\min_n\{p_n^1/p_n^0 : n = 1,...,N\} \le P(p^0,p^1) \le \max_n\{p_n^1/p_n^0 : n = 1,...,N\}$; i.e., the price index lies between the smallest and largest price relative.

T8: *Symmetric Treatment of Outlets*: $P(p^0,p^1) = P(p^{0*},p^{1*})$ where p^{0*} and p^{1*} denote the *same* permutation of the components of p^0 and p^1; i.e., if we change the ordering of the outlets (or households) from which we obtain the price quotations for the two periods, then the elementary index remains unchanged.

Eichhorn (1978; 155) showed that Tests 1, 2, 3 and 5 imply Test 7, so that not all of the above tests are logically independent.

[40] This was the approach used by Diewert (1995; 5-17), who drew on the earlier work of Eichhorn (1978; 152-160) and Dalén (1992).

The following tests are more controversial and are not necessarily accepted by all price statisticians.

T9: *The Price Bouncing Test*: $P(p^0,p^1) = P(p^{0*},p^{1**})$ where p^{0*} and p^{1**} denote possibly *different* permutations of the components of p^0 and p^1; i.e., if the ordering of the price quotes for both periods is changed in possibly different ways, then the elementary index remains unchanged.

Obviously, T8 is a special case of T9 where the two permutations of the initial ordering of the prices are restricted to be the same. Thus T9 implies T8. Test T9 is due to Dalén (1992; 138). He justified this test by suggesting that the price index should remain unchanged if outlet prices "bounce" in such a manner that the outlets are just exchanging prices with each other over the two periods. While this test has some intuitive appeal, it is not consistent with the idea that the price of a specific product in a specific outlet (which may have some special characteristics which are not present in other outlets) should be matched with the same product price in the same outlet in a one to one manner across the two periods.[41]

The following test was also proposed by Dalén (1992) in the elementary index context:

T10: *Time Reversal*: $P(p^1,p^0) = 1/P(p^0,p^1)$; i.e., if the data for periods 0 and 1 are interchanged, then the resulting price index should equal the reciprocal of the original price index.

It is difficult to accept an index that gives a different answer if the ordering of time is reversed.

T11: *Circularity*: $P(p^0,p^1)P(p^1,p^2) = P(p^0,p^2)$; i.e., the price index going from period 0 to 1 times the price index going from period 1 to 2 equals the price index going from period 0 to 2 directly.

The circularity and identity tests imply the time reversal test; (just set $p^2 = p^0$). The circularity property would seem to be a very desirable property: it is a generalization of a property that holds for a single price relative.

Elementary price indices may be calculated as direct price indexes by comparing the prices of the current period with those of a fixed price reference period or as chained short-term indexes obtained by multiplying the monthly (or quarterly) price indexes into a long-term price index. Many statistical offices chose to calculate the elementary price indices by chaining the short term (monthly or quarterly) indexes because this has some practical advantages when dealing with replacements in the sample. For elementary indexes calculated as chained shortcterm price indexes it is crucial that the index meets the circularity test.

T12: *Commensurability*: $P(\lambda_1p_1^0,..., \lambda_Np_N^0; \lambda_1p_1^1,..., \lambda_Np_N^1) = P(p_1^0,...,p_N^0; p_1^1,...,p_N^1) = P(p^0,p^1)$ for all $\lambda_1 > 0$, ..., $\lambda_N > 0$; i.e., if we change the units of measurement for each commodity in each outlet, then the elementary index remains unchanged.

In the bilateral index context, virtually every price statistician accepts the validity of this test. However, in the elementary context, this test is more controversial. If the N items in the

[41] Since a typical official Consumer Price Index consists of approximately 600 to 1000 separate strata where an elementary index needs to be constructed for each stratum, it can be seen that many strata will consist of quite heterogeneous items. Thus for a fruit category, some of the N items whose prices are used in the elementary index will correspond to quite different types of fruit with quite different prices. Randomly permuting these prices in periods 0 and 1 will lead to very odd price relatives in many cases, which may cause the overall index to behave badly unless the Jevons or Dutot formula is used.

elementary aggregate are all very homogeneous, then it makes sense to measure all of the items in the same units. Hence, if we change the unit of measurement in this homogeneous case, then test T12 should restrict all of the λ_n to be the same number (say λ) and test T12 becomes the following test:

(30) $P(\lambda p^0,\lambda p^1) = P(p^0,p^1)$ for all $p^0 >> 0_N$, $p^1 >> 0_N$ and $\lambda > 0$.

Note that (30) will be satisfied if tests T5 and T6 are satisfied.

However, in actual practice, elementary strata may not be very homogeneous: there may be thousands of individual items in each elementary aggregate and the hypothesis of item homogeneity may not be warranted. Under these circumstances, it is important that the elementary index satisfy the commensurability test, since the units of measurement of the heterogeneous items in the elementary aggregate are arbitrary and hence *the price statistician can change the index simply by changing the units of measurement for some of the items*.

This completes the listing of the tests for an elementary index. There remains the task of evaluating how many tests are passed by each of the five elementary indexes defined in section 2 above.

The following results hold:

- The Jevons elementary index P_J satisfies *all* of the above tests.
- The Dutot index P_D satisfies all of the tests with the important exception of the Commensurability Test T12, which it fails.
- The Carli and Harmonic elementary indexes, P_C and P_H, fail the price bouncing test T9, the time reversal test T10 and the circularity test T11 but pass the other tests.
- The geometric mean of the Carli and Harmonic elementary indexes, P_{CSWD}, fails only the (suspect) price bouncing test T9 and the circularity test T11.

Since the Jevons elementary index P_J satisfies *all* of the tests, it emerges as being "best" from the viewpoint of the axiomatic approach to elementary indexes.

The Dutot index P_D satisfies all of the tests with the important exception of the Commensurability Test T12, which it fails. If there are heterogeneous items in the elementary aggregate, this is a rather serious failure and hence price statisticians should be careful in using this index under these conditions. However, if the N items under consideration are all measured in the same units and the products are close substitutes, then the Dutot index could be used.[42]

The geometric mean of the Carli and Harmonic elementary indexes fail only the (suspect) price bouncing test T9 and the circularity test T11. The failure of test T9 is probably not a fatal failure and P_{CSWD} will usually be numerically to P_J so it will be close to satisfying the circularity test.

The Carli and Harmonic elementary indexes, P_C and P_H, fail the (suspect) price bouncing test T9, the time reversal test T10 and the circularity test T11 and pass the other tests. The failure of the time reversal test T10 (with an upward bias for the Carli and a downward bias for the Harmonic) is a rather serious failure and so price statisticians should not use these indexes.

---

[42] Evans (2012; 4) compared the Slovenian CPI with its corresponding Harmonized Index of Consumer Prices (HICP) and found very little difference over the period 1998-2011. The Slovenian national CPI used Dutot indexes at the elementary level and the Slovenian HICP used Jevons indexes at the elementary level.

In the following section, we present an argument due originally to Irving Fisher on why it is desirable for an index number formula to satisfy the time reversal test.

7. Fisher's Rectification Procedure and the Time Reversal Test

There is a problem with the Carli and Harmonic indexes that was first pointed out by Irving Fisher:[43] the rate of price change measured by the index number formula between two periods is dependent on which period is regarded as the base period. Thus the Carli index, $P_C(p^0,p^1)$ as defined by (2), takes period 0 as the base period and calculates (one plus) the rate of price change between periods 0 and 1.[44] Instead of choosing period 0 to be the base period, we could equally choose period 1 to be the base period and measure a reciprocal inflation rate going *backwards* from period 1 to period 0 and this *backwards measured inflation rate* would be $\sum_{n=1}^{N} (1/N)(p_n^0/p_n^1)$. In order to make this backwards inflation rate comparable to the forward inflation rate, we then take the reciprocal of $\sum_{n=1}^{N} (1/N)(p_n^0/p_n^1)$ and thus the overall inflation rate going from period 0 to 1 using period 1 as the base period is the following *Backwards Carli index* P_{BC}:[45]

$$(31)\ P_{BC}(p^0,p^1) \equiv [\textstyle\sum_{n=1}^{N} (1/N)(p_n^1/p_n^0)^{-1}]^{-1} = P_H(p^0,p^1) ;$$

i.e., the Backwards Carli index turns out to equal the Harmonic index $P_H(p^0,p^1)$ defined earlier by (4).

If the forward and backwards methods of computing price change between periods 0 and 1 using the Carli formula were equal, then we would have the following equality:[46]

$$(32)\ P_C(p^0,p^1) = P_H(p^0,p^1).$$

Fisher argued that a good index number formula should satisfy (32) since the end result of using the formula should not depend on which period was chosen as the base period.[47] This seems to be a persuasive argument: if for whatever reason, a particular formula is favoured, where the base period 0 is chosen to be the period that appears before the comparison period 1, then the same arguments that justify the forward looking version of the index number formula can be used to justify the backward looking version. If the forward and backward versions of the index agree

---

[43] "Just as the very idea of an index number implies a set of commodities, so it implies two (and only two) times (or places). Either one of the two times may be taken as the 'base'. Will it make a difference which is chosen? Certainly it *ought* not and our Test 1 demands that it shall not. More fully expressed, the test is that the formula for calculating an index number should be such that it will give the same ratio between one point of comparison and the other point, *no matter which of the two is taken as the base*." Irving Fisher (1922; 64).

[44] Instead of calculating price inflation between periods 0 and 1, period 1 can be replaced by any period t that follows period 1; i.e., p^1 in the Carli formula $P_C(p^0,p^1)$ can be replaced by p^t and then the index $P_C(p^0,p^t)$ measures price change between periods 0 and t. The arguments concerning $P_C(p^0,p^1)$ that follow apply equally well to $P_C(p^0,p^t)$.

[45] Fisher (1922; 118) termed the backward looking counterpart to the usual forward looking index the *time antithesis* of the original index number formula. Thus P_H is the time antithesis to P_C. The Harmonic index defined by (4) is also known as the Coggeshall (1887) index.

[46] Of course, equation (32) is *not* satisfied.

[47] "The justification for making this rule is twofold: (1) no reason can be assigned for choosing to reckon in one direction which does not also apply to the opposite, and (2) such reversibility does apply to any *individual* commodity. If sugar costs twice as much in 1918 as in 1913, then necessarily it costs half as much in 1913 as in 1918." Irving Fisher (1922; 64).

with one another, then it does not matter which version is used and this equality provides a powerful argument in favour of using the formula. If the two versions do not agree, then rather than picking the forward version over the backward version, a more symmetric procedure would be to take an average of the forward and backward looking versions of the index formula.

Fisher provided an alternative way for justifying the equality of the two indexes in equation (32). He argued that the forward looking inflation rate using the Carli formula is $P_C(p^0,p^1) = \sum_{n=1}^N (1/N)(p_n^1/p_n^0)$. As noted above, the backwards looking inflation rate using the Carli formula is $\sum_{n=1}^N (1/N)(p_n^0/p_n^1) = P_C(p^1,p^0)$. Fisher[48] argued that the product of the forward looking and backward looking indexes should equal unity; i.e., a good formula *should* satisfy the following equality (which is equivalent to (32)):

(33) $P_C(p^0,p^1)P_C(p^1,p^0) = 1.$

But (33) is the usual *time reversal test* that was listed in the previous section. Thus Fisher provided a reasonably compelling case for the satisfaction of this test.

As we have seen in section 4 above,[49] the problem with the Carli formula is that it not only does not satisfy the equalities (32) or (33) but it *fails* (33) with the following definite inequality:

(34) $P_C(p^0,p^1)P_C(p^1,p^0) > 1$

unless the price vector p^1 is proportional to p^0 (so that $p^1 = \lambda p^0$ for some scalar $\lambda > 0$), in which case, (33) will hold. The main implication of the inequality (34) is that the use of *the Carli index will tend to give higher measured rates of inflation* than a formula that satisfies the time reversal test (using the same data set and the same weighting).

Fisher showed how the downward bias in the backwards looking Carli index P_H and the upward bias in the forward looking Carli index P_C could be cured. The Fisher *time rectification procedure*[50] as a general procedure for obtaining a bilateral index number formula that satisfies the time reversal test works as follows. Given a bilateral price index P, Fisher (1922; 119) defined the *time antithesis* P° for P as follows:

(35) $P^{\circ}(p^0,p^1,q^0,q^1) \equiv 1/P(p^1,p^0,q^1,q^0).$

Thus P° is equal to the reciprocal of the price index that has reversed the role of time, $P(p^1,p^0,q^1,q^0)$. Fisher (1922; 140) then showed that the geometric mean of P and P°, say $P^* \equiv [P\times P^{\circ}]^{1/2}$, satisfies the time reversal test, $P^*(p^0,p^1,q^0,q^1)P^*(p^1,p^0,q^1,q^0) = 1$.

In the present context, P_C is only a function of p^0 and p^1, but the same rectification procedure works and the time antithesis of P_C is the harmonic index P_H. Applying the Fisher rectification procedure to the Carli index, the resulting rectified Carli formula, P_{RC}, turns out to equal the Carruthers, Sellwood and Ward (1980) and Dalén elementary index P_{CSWD} defined earlier by (5):

[48] "Putting it in still another way, more useful for practical purposes, the forward and backward index number multiplied together should give unity." Irving Fisher (1922; 64).

[49] Recall the inequalities (7) and (8) above.

[50] Actually, Walsh (1921b; 542) showed Fisher (1921) how to rectify a formula so it would satisfy the factor reversal test and Fisher (1922) simply adapted the methodology of Walsh to the problem of rectifying a formula so that it would satisfy the time reversal test.

(36) $P_{RC}(p^0,p^1) \equiv [P_C(p^0,p^1)P_{BC}(p^0,p^1)]^{1/2} = [P_C(p^0,p^1)P_H(p^0,p^1)]^{1/2} = P_{CSWD}(p^0,p^1)$.

Thus P_{CSWD} is the geometric mean of the forward looking Carli index P_C and its backward looking counterpart $P_{BC} = P_H$, and, of course, P_{CSWD} will satisfy the time reversal test.

8. Conclusion

The main results in this chapter can be summarized as follows:

- In order to define a "best" elementary index number formula, it is necessary to have a target index number concept. In section 2, it is suggested that normal bilateral index number theory applies at the elementary level as well as at higher levels and hence the target concept should be one of the Fisher, Törnqvist or Walsh formulae.
- When aggregating the prices of the same narrowly defined item within a period, the narrowly defined unit value is a reasonable target price concept.
- The axiomatic approach to traditional elementary indexes (i.e., no quantity or value weights are available) supports the use of the Jevons formula under most circumstances.[51] If the items in the elementary aggregate are very homogeneous (i.e., they have the same unit of measurement), then the Dutot formula could be used. In the case of a heterogeneous elementary aggregate (the usual case), the Carruthers, Sellwood and Ward formula can be used as an alternative to the Jevons formula but both will give much the same numerical answers.
- The Carli index has an upward bias (relative to the time reversal test) and the Harmonic index has a downward bias.
- All five unweighted elementary indexes are not really satisfactory. A much more satisfactory approach would be to collect quantity or value information along with price information and form sample superlative indexes as the preferred elementary indexes. However, if a chained superlative index is calculated, it should be examined for chain drift; i.e., a chained index should only be used if the data are relatively smooth and subject to long term trends rather than short term fluctuations.[52]

References

Balk, B.M. (1980), "A Method for Constructing Price Indices for Seasonal Commodities", *Journal of the Royal Statistical Society A* 143, 68-75.

Balk, B.M. (1994), "On the First Step in the Calculation of a Consumer Price Index", paper presented at the first Ottawa Group Meeting, Ottawa, October 31-November 4.

Balk, B.M. (2000), "On Curing the CPI's Substitution and New Goods Bias", Research Paper 0005, Department of Statistical Methods, Statistics Netherlands, Voorburg.

Balk, B.M. (2002), "Price Indexes for Elementary Aggregates: The Sampling Approach", Research Report, Voorburg: Statistics Netherlands.

---

[51] One exception to this advice is when a price can be zero in one period and positive in another comparison period. In this situation, the Jevons index will fail and the corresponding item will have to be ignored in the elementary index.

[52] If the price and quantity data are subject to large fluctuations, then multilateral methods should be used instead of a bilateral index number formula. Multilateral methods will be studied in chapter 7.

Balk, B.M. (2008), *Price and Quantity Index Numbers*, New York: Cambridge University Press.

Carli, Gian-Rinaldo, (1804), "Del valore e della proporzione de' metalli monetati", pp. 297-366 in *Scrittori classici italiani di economia politica*, Volume 13, Milano: G.G. Destefanis (originally published in 1764).

Carruthers, A.G., D.J. Sellwood and P.W. Ward (1980), "Recent Developments in the Retail Prices Index", *The Statistician* 29, 1-32.

Coggeshall, F. (1887), "The Arithmetic, Geometric and Harmonic Means", *Quarterly Journal of Economics* 1, 83-86.

Dalén, J. (1992), "Computing Elementary Aggregates in the Swedish Consumer Price Index," *Journal of Official Statistics* 8, 129-147.

Dalén, J. (1994), "Sensitivity Analyses for Harmonizing European Consumer Price Indices", pp. 147-171 in *International Conference on Price Indices: Papers and Final Report*, First Meeting of the International Working Group on Price Indices, November, Ottawa: Statistics Canada.

Davies, G.R. (1924), "The Problem of a Standard Index Number Formula", *Journal of the American Statistical Association* 19, 180-188.

Davies, G.R. (1932), "Index Numbers in Mathematical Economics", *Journal of the American Statistical Association* 27, 58-64.

de Haan, J. and F. Krsinich (2014), "Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes," *Journal of Business and Economic Statistics* 32, 341–358.

de Haan, J. and H. van der Grient (2011), "Eliminating Chain Drift in Price Indexes Based on Scanner Data", *Journal of Econometrics* 161, 36-46.

Diewert, W.E. (1978), "Superlative Index Numbers and Consistency in Aggregation", *Econometrica* 46, 883-900.

Diewert, W.E. (1995), "Axiomatic and Economic Approaches to Elementary Price Indexes", Discussion Paper No. 95-01, Department of Economics, University of British Columbia, Vancouver, Canada.

Diewert, W.E. (1998), "Index Number Issues in the Consumer Price Index", *Journal of Economic Perspectives* 12:1, 47-58.

Diewert, W.E. (2002), "Harmonized Indexes of Consumer Prices: Their Conceptual Foundations", *Swiss Journal of Economics and Statistics* 138:4, 547-637.

Diewert, W.E., J. de Haan and K.J. Fox (2016), "A Newly Identified Source of Potential CPI Bias: Weekly versus Monthly Unit Value Price Indexes", *Economics Letters* 141, 169-172.

Diewert, W.E. and P. von der Lippe (2010), “Notes on Unit Value Index Bias”, *Journal of Economics and Statistics* 230, 690-708.

Dutot, Charles, (1738), *Réflexions politiques sur les finances et le commerce,* Volume 1, La Haye: Les frères Vaillant et N. Prevost.

Eichhorn, W. (1978), *Functional Equations in Economics*, London: Addison-Wesley.

Evans, B. (2012), “International Comparison of the Formula Effect between the CPI and the RPI”, Office for National Statistics (ONS), Newport, UK.

Feenstra, R.C. (1994), “New Product Varieties and the Measurement of International Prices”, *American Economic Review* 34, 157-177.

Feenstra, R.C. and M.D. Shapiro (2003), “High Frequency Substitution and the Measurement of Price Indexes”, pp. 123-146 in *Scanner Data and Price Indexes*, Studies in Income and Wealth Volume 64, R.C. Feenstra and M.D. Shapiro (eds.), Chicago: The University of Chicago Press.

Fisher, I. (1921), “The Best Form of Index Number”, *Quarterly Publication of the American Statistical Association* 17, 533-537.

Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton-Mifflin.

Fox, K.J. and I. A Syed (2016), “Price Discounts and the Measurement of Inflation”, *Journal of Econometrics* 191, 398-406.

Gudnason, R. (2003), “How do we Measure Inflation? Some Measurement Problems”, 7th Ottawa Group Meeting on Price Indices, May 27-29, Paris.

Hardy, G.H., J.E. Littlewood and G. Pólya (1934), *Inequalities*, Cambridge: Cambridge University Press.

Hawkes, W.J. (1997), “Reconciliation of Consumer Price Index Trends in Average Prices for Quasi-Homogeneous Goods Using Scanning Data”, Third Meeting of the International Working Group on Price Indices, Statistics Netherlands, Voorburg, April 16-18.

Hawkes, W.J. and F.W. Piotrowski (2003), “Using Scanner Data to Improve the Quality of Measurement in the Consumer Price Index”, pp. 17-38 in *Scanner Data and Price Indexes*, Studies in Income and Wealth Volume 64, R.C. Feenstra and M.D. Shapiro (eds.), Chicago: The University of Chicago Press.

ILO IMF OECD UNECE Eurostat World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, Peter Hill (ed.), Geneva: International Labour Office.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), “Scanner Data, Time Aggregation and the Construction of Price Indexes”, *Journal of Econometrics* 161, 24-35.

Ivancic, L. and K.J. Fox (2013), “Understanding Price Variation Across Stores and Supermarket Chains: Some Implications for CPI Aggregation Methods”, *Review of Income and Wealth* 59:4, 629-647.

Jevons, W.S., (1865), "The Variation of Prices and the Value of the Currency since 1782", *Journal of the Statistical Society of London* 28, 294-320; reprinted in *Investigations in Currency and Finance* (1884), London: Macmillan and Co., 119-150.

Koskimäki, T. and M. Ylä-Jarkko (2003), "Segmented Markets and CPI Elementary Classifications", 7th Ottawa Group Meeting on Price Indices, Paris, May 27-29.

Koskimäki, T. and Y. Vartia (2001), "Beyond Matched Pairs and Griliches Type Hedonic Methods for Controlling Quality Changes in CPI Subindices", 6th Ottawa Group Meeting on Price Indices, Canberra, April 2-6.

Ley, E. (2003), "Comment", pp. 379-382 in *Scanner Data and Price Indexes*, Studies in Income and Wealth Volume 64, R.C. Feenstra and M.D. Shapiro (eds.), Chicago: The University of Chicago Press.

Reinsdorf, M. (1994), "The Effect of Price Dispersion on Cost of Living Indexes", *International Economic Review* 35, 137-149.

Reinsdorf, M. (1996), "Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee", BLS Working Paper 277, Bureau of Labor Statistics, Washington D.C., April.

Reinsdorf, M. and B.R. Moulton (1997), "The Construction of Basic Components of Cost of Living Indexes", pp. 397-436 in *The Economics of New Goods*, T. F. Bresnahan and R.J. Gordon (eds.), University of Chicago Press, Chicago.

Richardson, D.H. (2003), "Scanner Indexes for the Consumer Price Index", pp. 39-65 in *Scanner Data and Price Indexes*, Studies in Income and Wealth Volume 64, R.C. Feenstra and M.D. Shapiro (eds.), Chicago: The University of Chicago Press.

Silver, M. (1995), "Elementary Aggregates, Micro-Indices and Scanner Data: Some Issues in the Compilation of Consumer Price Indices", *Review of Income and Wealth* 41, 427-438.

Silver, M. (2010), "The Wrong and Rights of Unit Value Indices", *Review of Income and Wealth* 56:S1, 206-206.

Silver, M. (2011), "An Index Number Formula Problem: the Aggregation of Broadly Comparable Items", *Journal of Official Statistics* 27:4, 1–17.

Silver, M. and S. Heravi (2001), "Scanner Data and the Measurement of Inflation", *The Economic Journal*, 111, F384-F405.

Silver, M. and S. Heravi (2003), "The Measurement of Quality Adjusted Price Changes", pp. 277-316 in *Scanner Data and Price Indexes*, Studies in Income and Wealth Volume 64, R.C. Feenstra and M.D. Shapiro (eds.), Chicago: The University of Chicago Press.

Silver, M. and S. Heravi (2005), "A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment using Scanner Data", *Journal of Business and Economic Statistics* 23:3, 269–281.

Szulc, B.J. (1983), "Linking Price Index Numbers," pp. 537-566 *in Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

Szulc, B.J. (1987), "Price Indices below the Basic Aggregation Level", *Bulletin of Labour Statistics* 2, 9-16.

Theil, H. (1954), *Linear Aggregation of Economic Relations*, Amsterdam: North-Holland.

Triplett, J.E. (2003), "Using Scanner Data in Consumer Price Indexes: Some Neglected Conceptual Considerations", pp. 151-162 in *Scanner Data and Price Indexes*, Studies in Income and Wealth Volume 64, R.C. Feenstra and M.D. Shapiro (eds.), Chicago: The University of Chicago Press.

Triplett, J.E. (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, Directorate for Science, Technology and Industry, DSTI/DOC(2004)9, Paris: OECD.

Walsh, C.M. (1901), *The Measurement of General Exchange Value*, New York: Macmillan and Co.

Walsh, C.M. (1921a), *The Problem of Estimation*, London: P.S. King & Son.

Walsh, C.M. (1921b), "Discussion", *Journal of the American Statistical Association* 17, 537-544.

White, A.G. (1999), "Measurement Biases in Consumer Price Indexes", *International Statistical Review* 67:3, 301-325.

White, A.G. (2000), "Outlet Types and the Canadian Consumer Price Index", *Canadian Journal of Economics* 33, 488-505.